LASER OPTIMIZATION TECHNIQUES

C. D. Orth

S. W. Haney

W. H. Williams

Introduction

One of the challenges in designing a large laser system like the proposed National Ignition Facility (NIF) is determining which configuration of components meets the mission needs at the lowest construction cost with the largest margin in performance without any optical damage. Such an endeavor could credibly require evaluation of a million different designs because we must consider a range of values for each of many parameters: different numbers of beamlets, different aperture sizes, different numbers of laser slabs, different thicknesses of slabs, etc. The evaluation of so many designs with even a minimum of detail is impractical with current computing ability. There is hence a premium on the development of techniques to determine the best design without evaluating all of the practical configurations.

What Is an Optimization?

Optimization is the process of finding the best set of values for a system's identified design parameters. There are three steps to optimization. First, we identify exactly what aspects or measures of the system would define a better system if they were enhanced, such as improved output, lowered cost, improved reliability, or a combination of measures with appropriate weighting factors. This is often the most difficult step, because it requires that we truly understand what constitutes a better system. Second, we define a mathematical function known as a figure of merit (FOM) that incorporates these measures and has both of the following characteristics: (1) the value of the FOM increases or decreases according to our selected measure of goodness, and (2) the FOM is dependent on all of the design parameters under study, such that the value of the FOM varies if the value of any parameter is varied. Third, we search the parameter space to find the design that maximizes or minimizes the FOM, depending on how the FOM is defined.

To optimize a large laser system, we must consider more than just an FOM. Much of the physics of light propagation is nonlinear, in the sense that a change in each of the many parameters does not produce a proportional change in the laser output. This means that we cannot use any of the linear optimization methods that are known. There are also many equality and inequality constraints to consider, such as not exceeding a certain construction cost, not exceeding a certain input energy, not damaging any of the optical elements, and outputting at least a certain 3ω laser energy. Optimization of a laser system is thus not only nonlinear, but also constrained, and is hence called nonlinear constrained optimization. In addition, because a graph of the FOM plotted as a function of all of the variable parameters is often a multiple-peaked surface, with several peaks of different heights, there are local maxima (peaks) in the FOM that can easily fool a computer into incorrectly "thinking" it has found the optimum. What we need, therefore, is a method that can quickly and correctly find the highest peak in the FOM space, the so-called global optimum. Moreover, the FOM space is sometimes noisy in one or more of the parameters, with additional peaks and valleys superimposed on top of the otherwise smooth topology. We thus need a technique that can smooth out the noise. A mathematical statement of such a nonlinear constrained optimization problem is to maximize (or minimize) a suitably smoothed (or averaged) FOM function $f(\mathbf{x})$ subject to equality constraints $c_i(\mathbf{x}) = 0$, $i = 1, ..., k$, and inequality constraints $c_i(\mathbf{x}) \geq 0$, $i = k+1, ..., m$, where parameter vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ represents n variables.

Selecting an Optimization Technique

Many different techniques exist for nonlinear constrained optimization. These techniques differ in the way in which the computer navigates the FOM space in search of the tallest peak or deepest valley. The most common methods sample the topology in some region to determine the slope of the terrain in various directions, and hence the direction that leads to the optimum. These methods, which are all iterative, differ in the following ways in which the local terrain is sampled:

1. The downhill simplex method[1] uses only function evaluations (i.e., evaluations of the FOM), not derivatives, to move in the direction of the optimum using manipulations of the geometrical figure (simplex) formed by the original point and n somewhat arbitrary displacements in the n directions corresponding to the n parameters considered. This method is robust, but typically slow to converge.
2. Simple derivative (gradient) methods sample the space for two (or more) values of each parameter to determine a local slope in the FOM, and then step off a fixed step in the direction of maximum slope. One serious problem with this method is that, once the best direction in which to go for the next calculation is identified, we don't really know how far to go in that direction. Consequently, this method can be slow to converge.
3. Quadratic methods, in essence, calculate the FOM for three values of each parameter, fit the resulting FOM values to a quadratic function such as a parabola, and calculate exactly how far to go for the next estimate of where the FOM peak might lie. This method is much "smarter" than the above methods because it knows how far to step out for the next iteration. Two common methods of this type are Brent's method and Powell's method. Brent's method uses parabolic interpolation. Powell's method finds the solution that zeros the gradient of a Lagrange function formed as a sum of the FOM function and a linear combination of the constraint functions. It approximates the full nonlinear problem by a quadratic problem to find the direction to go in the FOM space, then determines how far to go in that direction by doing a one-dimensional optimization in that direction. While Brent's method uses only three function evaluations for each parameter, Powell's method uses one function evaluation, one derivative evaluation, one estimate for the second derivative, plus other function evaluations for each subsequent line search. These methods are powerful and fast, and are the methods used in the codes that we employ.
4. Simulated annealing methods[2] randomly sample points over a large portion of the parameter space and then slowly decrease the size of the sampling space around any optimum so found, according to mathematics analogous to Boltzmann thermodynamic cooling. This method is good for selecting the best global ordering of a large combination of discrete elements of some kind when many local optima exist, but it converges too slowly and is thus not practical for our problem.
5. Genetic methods[3] form a pool of design candidates, "mate" the ones having the highest FOMs (the parents) by combining attributes in some way to produce other candidates (the offspring), and then remove the unlikely candidates (extinction). This method can also be effective in finding a global optimum among many local optima, but it also converges too slowly.

The first three methods listed above can be confused if the FOM (or constraint) surface is noisy, unless additional smoothing algorithms are included. Simulated annealing and genetic methods are inherently more able to navigate successfully along noisy surfaces, but are expected to converge too slowly to be effective for the optimization of a large laser system. We therefore chose to use quadratic optimization methods to speed up the computational time. As a result, however, we had to deal with discontinuities in the FOM as one form of noise.

Dealing with Discontinuities in the FOM

There are several sources of fluctuations in the FOM that affect an optimizer evaluating various designs of a large laser system. Although stochastic (statistical) noise can sometimes be a problem when methods employ random numbers, a more typical problem arises from the discontinuous FOM jumps that correspond to integer steps. Integers are encountered in such parameters as the numbers of laser slabs in the amplifiers, the number of flashlamps for each slab, and the number of beamlets in the laser. If the FOM (or a constraint) moves abruptly as the optimizer searches from one integer to the next, as it usually does, the optimizer can be confused by the local optima thereby introduced. This is especially true for gradient-search techniques, which are based on real numbers. We can overcome such fluctuations by incorporating either some type of smoothing technique or by substituting real numbers for the integer quantities so that the optimizer can consider noninteger (i.e., nonphysical) values of the parameters.

A second source of discontinuities arises in optimizations involving spatial-filter pinholes because of the integer-like effect of the transverse spatial representation of the beam by a fixed number of grid points (e.g., 512 points over a horizontal width of 50 cm). To explain these discontinuities, we must first explain the purpose and function of pinholes.

All of the large laser chains at LLNL involve the use of spatial filters that focus the beam and force it to go through a 100- to 200-μm-diam pinhole in a metal plate. The purpose of such a pinhole is to cut off the high-angle noise present in the beam, thereby smoothing the transverse spatial intensity profile. Such smoothing occurs because, in the focal plane of a lens, the spatial intensity profile corresponds to the power in the various angular components in the beam. Thus, if the intensity profile is decomposed into its Fourier components, the relative powers in the low-frequency components correspond to the intensities seen near the centerline of the focal plane. Higher-frequency components focus at increasingly farther distances from the centerline because they are propagating at larger angles. These various frequency components grow in magnitude at different rates due to, among other things, the nonlinear indices of refraction of the optical materials in the laser, as described by the Bespalov–Talanov theory.[4] The high-frequency components tend to grow faster, so it becomes desirable to clip off the power in these frequencies periodically in the laser chain to keep the beam as spatially flat as possible. For a beam with a reasonably smooth spatial profile, the power in the high-frequency components is low enough to permit a pinhole to do this clipping without ablating away too much of the pinhole structure and thereby disrupting the last temporal portions of the laser pulse.

In propagation simulations employing a finite number of grid points, the power in the Fourier components is summed into the same number of bins as there are grid points. Consequently, as the diameter of the pinhole changes with changing pinhole acceptance angle, there are discrete jumps in clipped-beam power as the edge of the pinhole moves from one frequency bin (grid point) to the next. These jumps result in abrupt changes in the spatial intensity profile of the beam after it leaves the filter. If the optimizer is monitoring, for example, the peak fluence as a constraint against optical damage, it can be confused by the local optima that such discontinuities generate. The solution we have used to overcome this confusion is to average the FOM over several adjacent function evaluations.

Optimizing the NIF Design

Optimization of the NIF design is one example of how we can use our optimization techniques. As described in more detail in the article "NIF Design Optimization" on p. 181 of this *Quarterly*, the conceptual design[5] for the NIF configuration arose through use of a code called CHAINOP,[6] which was capable of using either a simplex optimization routine or a gradient-search routine. We accessed this gradient-search routine as part of a large systems analysis programming shell called SUPERCODE.[7] SUPERCODE uses the quadratic version of Powell's method, as written in an Argonne National Laboratory implementation of a variable metric method labeled VMCON.[8] Because VMCON evaluates gradients, care was taken to minimize discontinuities in the FOM and constraints by using noninteger laser slab and flashlamp counts, as well as smoothed cost functions.

In the second optimization effort, for the Advanced Conceptual Design (ACD) phase of the NIF, we implemented three separate approaches. In one approach, we used SUPERCODE to evaluate ~100,000 specific designs distributed around the conceptual design point in the major parameters, and we applied various cuts to constrain dynamical and engineering quantities such as cost and nonlinear growth of spatial noise. For a second approach, we used SUPERCODE to do an optimization. As with CHAINOP, we used fractional slabs and noninteger flashlamp counts, but optimization over pinhole acceptance angle could not be performed. A third approach, coded as OPTIMA1, is based on a modified version of Brent's method (parabolic interpolation). A similar approach had been used in the diode-pumped solid-state laser (DPSSL) design studies for inertial fusion energy.[9] OPTIMA1 can use either of two different methods to smooth the FOM space and can treat integer parameters directly. This code therefore allowed the consideration of noisy parameters, including integer slab and flashlamp counts, and pinhole acceptance angles.

There were several reasons we chose multiple approaches in optimizing the NIF laser for the ACD. First, when we started, we knew that there were complexities such as noisy parameters, and we were not sure that any one approach would prove successful in the end. Second, the performance of optimizers is usually dependent on the problem considered, so the length of time needed for any one optimizer to converge to a solution was unknown for the NIF problem. In general, performance is better if the optimization algorithm is tailored to the problem in some way. On the other hand, the use of a well tested fast algorithm like VMCON can often prove very effective. Third, we wanted multiple opinions from independent approaches to improve the reliability of the results.

All three approaches used parallel processing through Parallel Virtual Machine (PVM) software. This software allows one master computer to send tasks to many different slave computers and retrieve the results. In our case, we had one HP 715/80 UNIX master workstation tasking 28 similar slave workstations that each evaluated the performance of one NIF design (SUPERCODE) or performed the suboptimization for one parameter (OPTIMA1). The total computing power was 20–30 megaflops for each of 28 computers, or about 0.75 gigaflops. The use of this workstation cluster enabled the codes to search through a space representing a million candidate designs at a rate over 20 times faster than it would have taken with only one computer. This facility proved valuable because it reduced the total calculation time from the 46 days it would have taken for 10^6 assessments at about two minutes per assessment of each candidate design on one computer to about one or two days using the workstation cluster.

Features of SUPERCODE

SUPERCODE[7] couples a toolbox of systems analysis tools to a powerful user interface. The systems analysis tools currently consist of the following:

- A constrained nonlinear optimization package based on VMCON.[8]
- A nonlinear optimization package based on a genetic algorithm.[3]
- A nonlinear equation solver.
- A Monte Carlo sampling package for performing uncertainty analyses.
- Parameter scans.
- Uncertainty analyses.
- Facilities for exploiting PVM software to speed up optimizations.

These tools are controlled by a programmable shell that understands a large subset of the C++ computer language, including arithmetic expressions, loops, decision structures, functions, classes, and objects. Users can interactively manipulate the set of equations to be passed to the systems analysis tools, monitor progress of a calculation, postprocess results, and plot them using an interactive graphics package. SUPERCODE was the main code used in setting the NIF baseline design, and has been used to design tokamak reactors and experiments, hybrid electric vehicles, rail guns, and neutron sources. SUPERCODE is available on UNIX, MacOS, and Windows platforms.

The OPTIMA1 Code

OPTIMA1 is based on the particular parabolic-interpolation method that proved successful for the DPSSL inertial fusion energy study.[9] This method increases the independence of the parameters, often permitting the full optimization to be accomplished by independent one-parameter optimizations. The key to this method is to incorporate an inner loop that maximizes the laser's injection energy subject to all constraints. An outer loop then deals with all of the other parameters designated as active optimization variables. Both loops advance by making one function evaluation and using that together with the last two calculations nearest the current best value to determine a next guess. Penalty functions reduce the value of the FOM by exponentially degrading the FOM as any active variable violates a constraint. The current list of features includes the following:

- Easy addition of parameters.
- Treatment of both discrete (integer, even-integer, or odd-integer) and analog (real) parameters.
- Two types of smoothing for noisy parameters (random-number smoothing and perturbation smoothing).
- Ability to restart runs from a dump file.
- Operation on either a single or multiple (parallel-processing) platforms (using PVM software).
- Table-like scans over a single parameter.
- Input-file selection of the optimization FOM from among any product or quotient of the parameters or calculated quantities.
- Performance of a sensitivity study around the final parameter vector.

OPTIMA1 is very flexible and robust. It usually takes more time to converge to a solution for small ($<10^{-3}$) precisions, however, because the penalty functions create a "cliff" in the FOM space that requires special treatment in the parabolic interpolator. In addition to helping choose the NIF design parameters, OPTIMA1 has been useful in performing sensitivity studies on the parameters (e.g., the maximum allowable doubler detuning angle) and in determining optical specifications through studies varying the aberrations placed on the optical components. OPTIMA1 is written in UNIX HP FORTRAN.

Overcoming Specific Modeling Difficulties

After selecting a set of optimization techniques, we had to deal with many problems that were specific to the optimization of a large laser. First, we found many designs having the same value of energy delivered to the laser entrance hole (LEH) of the target, but with differing construction costs, so we had to include cost in the FOM or cost as a constraint. Two solutions that proved effective were (1) minimizing cost subject to performance constraints and (2) maximizing the ratio of energy delivered to the LEH and total cost, subject to the other constraints.

The second problem was finding a way to include realistic aberration sources for every optic that would account for surface roughness and bulk phase retardations. The challenge was to develop techniques to use measurements taken over varying aperture sizes and on a limited number of parts while preserving the spatial distributions of wavefront roughness in such a form that realistic aberration sources could be simulated for every optic. This was done by devising a technique using a power spectral density (PSD) function defined as the square of the discrete Fourier transform of the measured phase retardations, multiplied by the length of the 1–D measurement (or multiplied by the length and width in 2D).[10] We combined the measurements from three spatial-frequency ranges (0 to 0.1 mm^{-1}, 0.1 to 1 mm^{-1}, and 1 to 10 mm^{-1}) while mitigating the effects of finite sampling apertures by using a windowing filter function (a Hanning filter). This filter broadened (or blurred) the resulting width of the spatial frequency structures and eliminated frequencies generated by the discontinuities at the edges of the samples. The PSDs thereby obtained did not possess any relative phase information in the Fourier domain, but were merely the spectral distributions of the wavefront errors. We then used the PSD for a given type of optic to determine the magnitude of the Fourier coefficient and added a random phase in the Fourier domain. Using inverse Fourier transformation, we obtained a unique wavefront distortion for every optic needing a simulated phase-front aberration.

The third modeling difficulty was that laser gain and pump-induced distortions depend on the choice of pumping parameters (pump pulse duration, Nd doping concentration, flashlamp explosion fraction, flashlamp diameter, and flashlamp packing fraction). Pump-induced distortion is phase noise added to the beam because the nonuniform distribution of flashlamp light absorbed by the laser slabs causes nonuniform heating, which distorts the slabs. Because determining the proper pump-induced distortions is calculationally intensive (involving ray tracing and 3-D thermomechanical modeling), we ran the 3-D codes for a range of each of the pumping parameters and incorporated the results in the optimizations using a table look-up procedure.

The fourth problem was that the use of spatial-filter pinholes with abrupt edges in calculations with a finite number of transverse spatial grid points causes ringing (aliasing), which introduces artificial modulation of the downstream beam intensity profile. We therefore used high-resolution runs with many points to define problem areas and incorporated smoothing algorithms developed at LLNL.

The fifth problem was damage to the optical components that can arise from either high fluence or filamentation, which both depend on the noise on the beam. (Filamentation is the process by which the intensity of a narrow beamlet increases as the beamlet is focused by nonlinear propagation through solid materials.) A proper treatment of these effects therefore requires very high spatial resolution to define the peaks in the transverse intensity profiles. Such resolution would make the iteration cycle time prohibitively slow if many points had to be used in two transverse dimensions. Consequently, we performed most of our calculations using 1-D codes with 512 spatial points, which is insufficient to determine filamentation, and evaluated the likelihood of filament formation and the resultant optical damage using a phenomenological model[11] developed at LLNL based on measured Beamlet data. For ordinary (nonfilamentation) optical damage assessment, we used measured peak fluence limits as constraints for every component.

The sixth modeling difficulty was that the desired NIF laser pulse at the LEH is a 3ω pulse consisting of a long low-power foot followed by a main pulse roughly 3.5 ns wide. Because the laser input is at 1ω, and because the harmonic conversion from 1ω to 3ω is intensity-dependent, we had to incorporate precise algorithms for this conversion process (see the article entitled “Frequency Conversion Modeling” on p. 199 of this *Quarterly*); we also had to iterate to make sure that the 1ω pulse shape produced the desired output 3ω pulse shape.

Optimization Flow

Each optimization iteration included a number of steps involving other codes, as outlined in Fig.1 for the particular case of OPTIMA1 using the Ethernet connections to the cluster of 28 workstations. As directed by the master, a slave machine would suboptimize one parameter by first running two codes to help establish the temporal shape of the input beam: a plane-wave frequency conversion code (thgft02) to assess the converter performance with the given parameter values and an inverse harmonic-conversion code (invconv3) to calculate the 1ω input temporal shape that would give the desired 3ω output temporal shape with that converter. Following the formation of the apodized transverse spatial shape of the input beam, the optimizer would generate the input file for PROP92 and run PROP92 to propagate the input signal down the chain of optical components from the input of the laser to the harmonic converter. This step included phase aberration sources to simulate the experimental surface finishes and bulk properties for each optic, as well as the amplifier gain files generated from 2.5-D amplifier modeling and the pump-induced distortion files generated from 3-D thermomechanical modeling (see the companion article entitled “The PROP92 Fourier Beam Propagation Code” for a complete description of the physics in PROP92). After running a code to analyze the PROP92 output file, the optimizer would run a frequency conversion code (thgxtz001) to convert the 1ω light to 3ω light. Thereafter, the optimizer would formulate an input file for PROP92, this time for the 3ω light, and run PROP92 to propagate the beam down the rest of the laser chain to obtain the laser pulse entering the target LEH. After

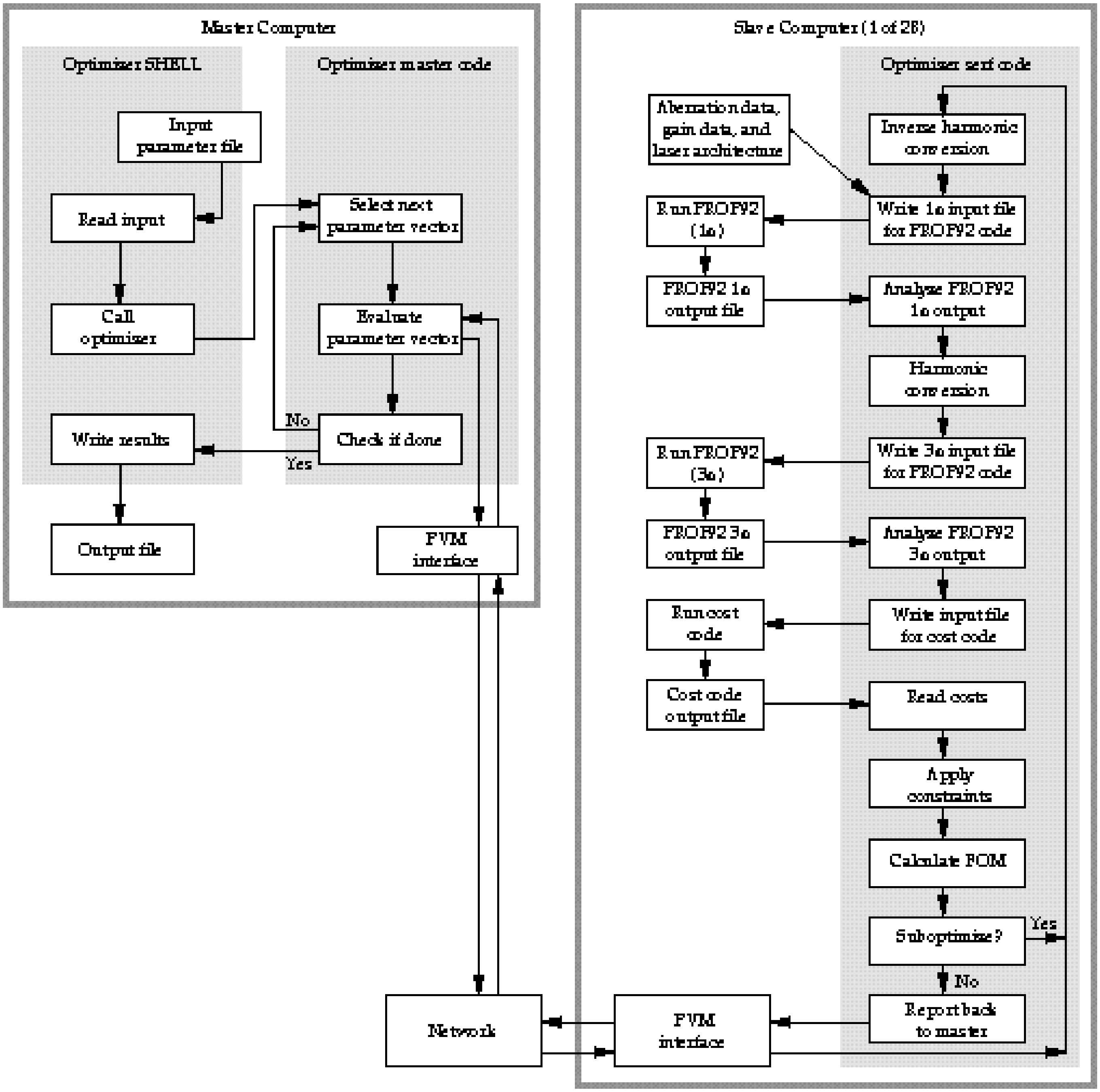

FIGURE 1. NIF optimization flow chart for OPTIMA1. (40-00-1296-2752pb01)

running a code to analyze the PROP92 3ω output file, the optimizer would run a routine to generate an input file for the cost code, run the cost code to calculate the resulting cost of the NIF facility, and run a routine to apply the constraints.The optimizer would then calculate the FOM as the LEH energy, or the ratio of that energy to the cost, or simply as the cost constrained to obtain the desired energy and power at the LEH—all as penalized by the constraints that failed. After several suboptimizations, the slave reports the results for that parameter back to the master, which decides how to form the full parameter vector for the next set of suboptimizations. When the changes in all parameters are less than specified precisions, the process stops.

Summary

Nonlinear constrained optimization of the design of a large laser such as the NIF is the process of finding the set of values for the laser's identified design parameters that optimizes the figure of merit (FOM). Among the different optimization techniques available, we selected quadratic methods based on Brent's method and Powell's method, as modified to treat the discontinuities imposed by integer parameter values. An established code (SUPERCODE) and a new code (OPTIMA1) were configured with realistic aberration sources for every optic, a cost model of the whole laser system, methods to incorporate the effects of different values of the pumping parameters, models for optical damage as well as filamentation, and full harmonic conversion of the desired 3ω pulse shape. We also incorporated parallel processing through use of Parallel Virtual Machine (PVM) software operating on a group of 28 workstations. The resulting techniques allow optimization of the NIF laser and other laser systems based on realistic components and realistic laser light propagation.

Notes and References

1. J. A. Nelder and R. Mead, *Computer Journal* **7**, 308–313 (1965).
2. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671–680 (1983); S. Kirkpatrick, *Journal of Statistical Physics* **34**, 975–986 (1984).
3. E.g., E. Betensky, *Opt. Eng*. **32,** 1750 (1993).
4. J. Trenholme, *Laser Program Annual Report*, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-50021-74, p. 177 (1975).
5. *National Ignition Facility Conceptual Design Report*, Lawrence Livermore National Laboratory, Livermore, CA, UCRL-PROP-117093 Vol. 2, NIF-LLNL-94-113 (1994).
6. J. Trenholme, "CHAINOP: Laser Chain Optimization Code," Lawrence Livermore National Laboratory, Livermore, CA, internal report (Oct. 1992); also see accompanying article on p. 181 in this *Quarterly*.
7. S. W. Haney, W. L. Barr, J. A. Crotinger, L. J. Perkins, et al., *Fusion Technology* **21**, 1749 (1992).
8. R. L. Crane, K. E. Hillstrom, and M. Minkoff, "Solution of the General Nonlinear Programming Problem with Subroutine VMCON," Report ANL-80-64, Argonne National Laboratory (1980).
9. C. D. Orth, S. A. Payne, and W. F. Krupke, *Nuclear Fusion* **36**(1), 75 (1996).
10. J. K. Lawson, D. M. Aikens, R. E. English, Jr., C. R. Wolfe, "Specification, Production, and Testing of Optical Components and Systems," SPEIE Proceedings, 13–16 May 1996, Glasgow, Scotland, UK, vol. 2775, p. 345 (1996).
11. Mark A. Henesian, "Simulations of 3ω Beam Filamentation in the Beamlet Focus Lens and General Comments on Filamentation Theory," Lawrence Livermore National Laboratory, Livermore, CA, internal memo LS&T-LMO 96-001 (May 1996); and "Simulations of 1ω Beam Filamentation in the Beamlet Laser, Filamentation Theory in Glass Slabs, and PROP92's Filamentation Code," Lawrence Livermore National Laboratory, Livermore, CA, internal memo, LS&T-LMO 96-003 (May 1996).